

# Context-Aware Retrieval-based Deep Commit Message Generation

Lorenzo Borelli, Simone Mattioli, Giovanni Minelli



# Objectives

- Reproduce CoRec model and results;
- Introduce Transformers;
- Use scheduled sampling and retrieval module with Transformers;



# Automatic commit message generation

- Produce commit messages starting from git diffs;
- 3 main techniques:
  - **Rule-based;**
  - **Retrieval-based;**
  - **Learning-based;**



# NMT

Learn optimal parameters to translate source sequence into a target language.



Encoder-decoder

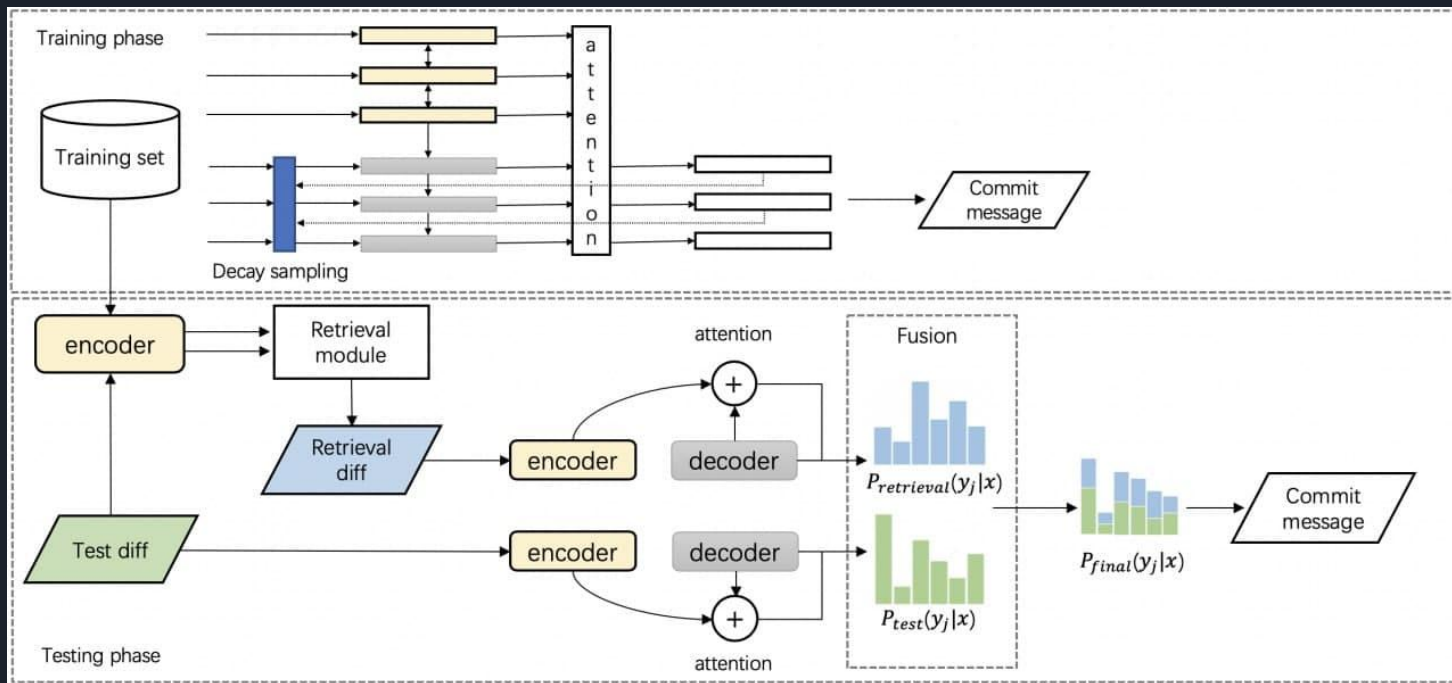


Exposure  
bias



Low frequency  
words

# CoRec



Preprocessed data  Train  Translate

# Transformer

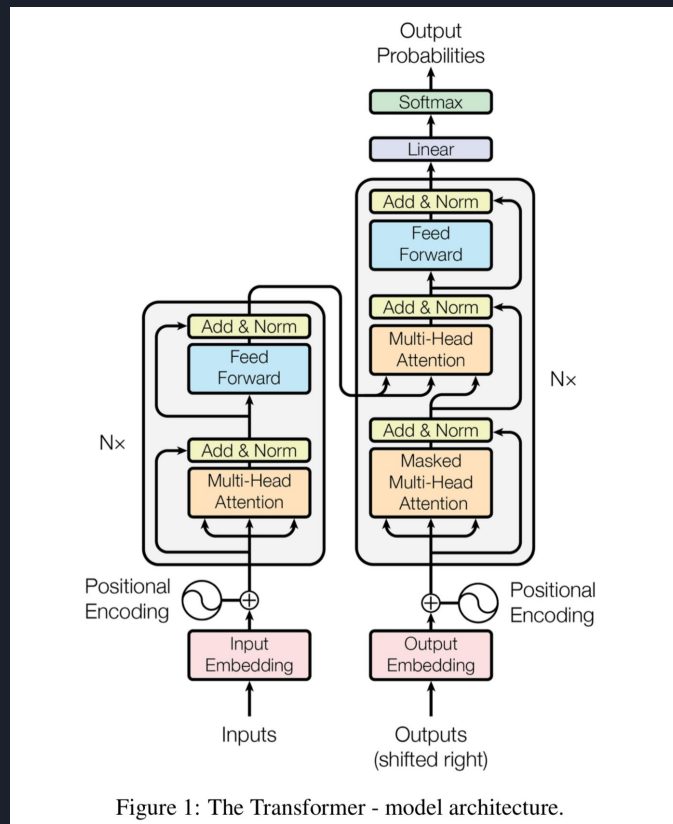
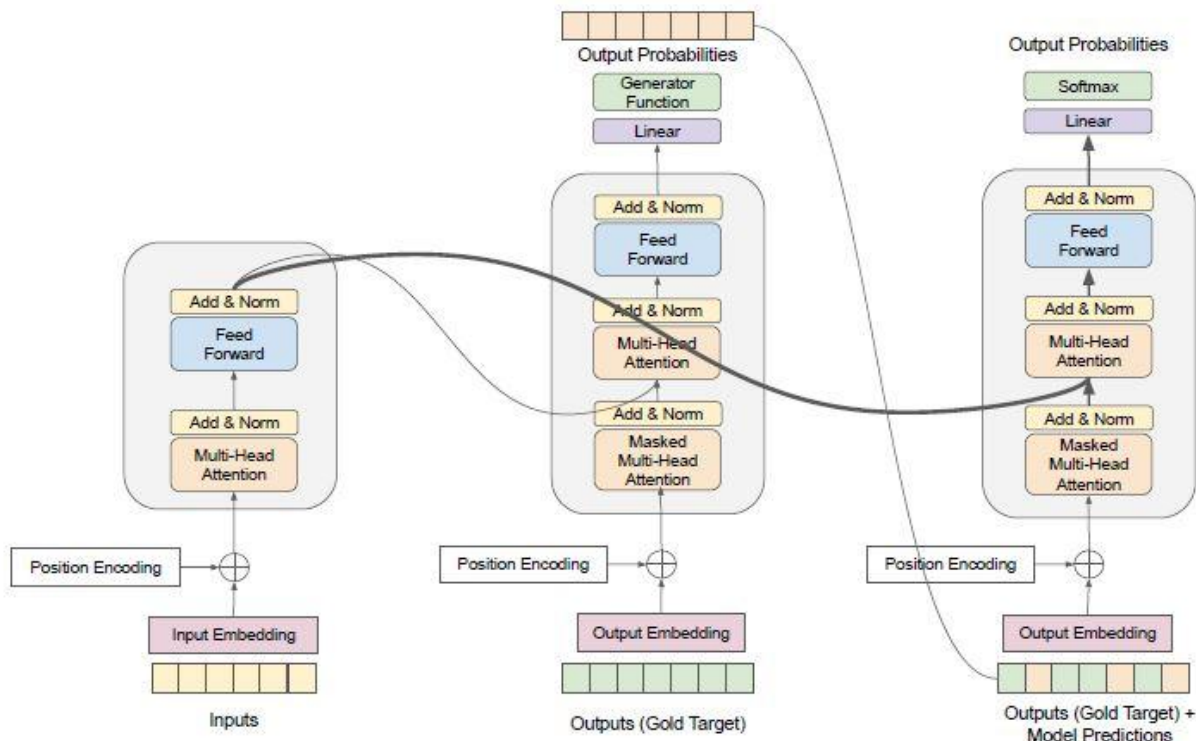


Figure 1: The Transformer - model architecture.

# Scheduled sampling






# Metrics

- Bleu
- Meteor
- Rouge-L
- BERTScore





## Results: top1000

Top 1000 repositories: 22K valid samples.



Low results  
respect LSTM



Overfitting with  
Transformers



# Results: top10000

Top 10000 repositories: 98K valid samples.

	Bleu	Rouge-L	Meteor	Bert F1
BiLSTM+ss+retrieval	<b>54.70%</b>	42.86%	39.88%	39.96%
Transformer	52.16%	43.51%	40%	40.06%
Transformer+retrieval	53.29%	<b>43.84%</b>	<b>40.19%</b>	<b>40.75%</b>
Transformer+ss	48.64%	42.25%	37.47%	37.87%
Transformer+ss+retrieval	50.79%	43.01%	38.84%	39.09%

Best results without dropout and scheduled sampling decay, but with retrieval module



# Output descriptions and errors

- Bias of Bleu
- High reliability of BertScore

Example

**Prediction:** modified ognlvariableexpressionevaluator constructor to protected .  
**Gold:** modified ognlvariableexpressionevaluator constructor to protected .



Generation of copied  
words out of context



Problems with  
long sentences



# Scheduled Sampling effects

Evidences from training and evaluation:

- Lower accuracy -
- Lower cross entropy and perplexity +
- Lower translation metrics -

From correlation analysis:

- All translation metrics become highly correlated
- Reduced correlation of translation metrics with training signals due to stochastic factor



# Conclusions

~ Transformers

✗ Scheduled sampling

✓ Retrieval module

? Attention role in the task



Project extension Attention analysis

? Other approaches



Project extension Task survey



# Attention role

Try to direct the focus of a network between all words regardless of their respective position in a sentence

- [Distribution analysis](#)
- Exploitation for difficulty reduction (Luong 2015)

(Transformer + retrieval)	Parameters	Bleu	Rouge-L	Meteor	Bert F1
Baseline	126235220	53.29%	43.84%	40.19%	40.75%
Targets reduced	102137983(-19.1%)	48.69%	41.06%	36.60%	34.19%
Pos_unk reduction	104339118 (-17.3%)	50.8%	41.73%	37.67%	36.67%



# Methods for automatic code summarization

## Rule-based

- Stereotypes
- Templates defined by devs
- Poor generalization

## Retrieval-based

- Description-code pairs
- Relevant information
- Too dependent on similarity

## Learning-based

- Semantic relation between source and target.
- Feature engineering



# NMT papers

Loyola et al.	Liu et al.	Siyuan et al.	Hu et al.	Tang et al.	Nie et al.	Tae-Hwan Jung
First NMT based approach to code summarization	Deal with OOV	VDO filter. QA with human evaluation.	AST, SBT, type annotation of OOV.	Relation matrices, tree attention.	Transformer for contextualized token embeddings.	Pre-trained CodeBERT for add/delete pairs.





# Conclusions

- NMT approaches are state-of-the-art.
- Some employ mixed approaches to handle domain-specific problems.
- Human judgement and evaluation is still needed.